

- 
- 1        6. The automated method for setting up a natural language interface in a  
2        Web site recited in claim 1, further comprising the step of converting the  
3        set of n-grams to classification rules.
- 

#### REMARKS

Acceptance of the informal drawings filed with the patent application for purposes of examination is noted with appreciation. Formal drawings will be submitted at such time as the application is allowed.

The title of the invention has been amended to delete the word "CONVERSATIONAL". On reconsideration, the word, which has its origins in an early IBM operating system, the Conversational Monitor System or CMS. As used herein, the term "conversational" generally means a program or system that carries on a dialog with a user. Perhaps a more current term would be "query system"; however, it is believed that this is clearly implied from the disclosure as filed. Similar amendments have been made to the specification and claims.

The specification also has been amended to correct a grammatical and a spelling error. No new matter has been added.

Claims 1 to 6 now appear in the application. Original claims 1 to 5 have been amended, and new claim 6 has been added by this amendment.

Claims 1 and 3 were rejected under 35 U.S.C. §102(e) as being anticipated by U.S. Patent No. 6,311,182 to Colbath et al. This rejection is respectfully traversed for the reason that Colbath et al. neither shows nor suggests the claimed invention.

The present invention provides an automated method for setting up a Web site with a natural language interface. The present invention is not directed toward speech recognition (although the present invention can be used in combination with speech recognition). With reference to Figure 2 of the drawings, in the present method, as claimed, a Web crawler 21, or similar program, creates a hierarchy of topics 22 from the Uniform Resource Locators (URLs) in a Web site

(see page 3, lines 5–14, and page 5, lines 19–22, of the present specification). Then, text on each page is analyzed to generate a keyword index 23; each node has an associated collection of selected keywords. These keywords can be n-grams, for example. The use of stochastic n-gram (Markovian) models has a long and successful history in the support of vocabulary applications in speech recognition systems. Applicants, however, use n-grams in a different way. The logic is as follows. Each topic has a set of n-grams, perhaps sparse, associated with it. Each (sparse) n-gram is connected to a rule in which each term of the n-gram is a term of a rule whose consequent is the topic associated with the n-gram being converted. As used herein, and in the specification, “n-gram” includes sparse n-gram and non-sparse n-gram. The distinction is made on page 3, lines 17–22. Formally, since a sparse n-gram is a set of ordered words (or tokens, etc.) within a window  $d$ , the traditional notion of an n-gram as a sequence of  $n$  words, is simply a sparse n-gram with  $d=n$ ; i.e., the length of the sequence with no gaps. In Applicants’ usage of n-grams, gaps are allowed between words in their n-grams, hence their n-grams can be sparse or not sparse. As noted in the specification on page 5, lines 7–8 and lines 11–13, the selection criterion can be the chi-square measure, or a statistical test confidence measure. In a final step, a mechanism 25 is specified for associating classification rules to the topic. Classification rules are created from the keywords or n-grams. For example, given the n-gram “need car loan”, which is statistically associated with the topic “car\_loan”, the rule “need & car & loan  $\rightarrow$  car\_loan” can be produced. This rule can be associated with topics relating to cars or loans.

Accordingly, the present invention provides an automated method for establishing a query interface for a Web site. The query interface allows for rapid and efficient searching of a Web site. The present invention does not necessarily make use of speech recognition, but speech recognition may be used in combination with the query interface provided by the invention.

The patent to Colbath et al., by comparison, teaches a very different technology; specifically, a voice-activated Web browser. In Colbath et al., voice

signals are recognized and converted into words. These words are used to form a search string, and a search is then performed, for example, on the Internet or on a website. The search is performed over a preselected collection of areas of interest. Colbath et al. further disclose methods for searching when the search terms do not match with any preselected areas of interest.

Colbath et al. is very different from the present invention as claimed for several reasons. First, the present invention is directed to a method for setting up a Web site query interface, and Colbath et al., by contrast, is directed towards searching based on voice commands. Colbath et al. do not teach setting up a Web query interface. Second, Colbath et al. do not teach the step of, for each website topic, associating a set of n-grams to the topic, which are distinctive of that topic, as recited in the third step of claim 1, as amended. In the preferred embodiment, these sets of n-grams are converted to classification rules, and new claim 6, dependent on claim 1, has been added to recite this step.

Colbath et al. do not teach or suggest an automatic method for setting up a Web query interface. In fact, Colbath et al. is completely lacking any suggestion to set up a query interface. Instead, Colbath et al. teaches only methods for conducting web searches using voice commands.

By comparison, independent claim 1 and dependent claim 3 are directed to "setting up a natural language interface in a Web site". Setting up a natural language interface according to the present invention requires that documents on a Web site are classified, and requires that a keyword index is created for documents in the Web site. This allows a person creating the natural language interface to do so efficiently and easily. The natural language interface allows a search engine to find documents on a Web site set up according to the invention. Colbath et al., do not teach how to create or set up a natural language interface, but instead teach how to perform a search using voice commands. Setting up a natural language interface and performing a search are two different and distinct functions. Setting up a natural language interface allows a search program to search a Web site according to a query protocol (possibly specified by the interface), and performing

a search finds documents of interest. Hence, the teachings of Colbath et al. are not really applicable to the claimed invention.

Specifically, because Colbath et al. do not teach setting up a natural language interface, and instead teach performing a search, they necessarily lack the essential step of “generating a keyword index for those documents”, as recited in claim 1. The Examiner argues that Colbath et al. teach this limitation in col. 3, lines 1–12. However, in this passage, Colbath et al. explain something quite different; specifically, that it is the “most probable word strings” of the *input speech* that are searched for. By comparison, in the present invention, the above-referenced limitation requires that a keyword index is *created for a collection of documents* so that the documents can be searched more effectively. The keyword index of the present invention allows a search engine to find documents; *the keyword index is not searched for*, as required by Colbath et al. Instead, the keyword index of the present invention *represents a field searched in*. The Examiner has confused the search terms with the search field in the Colbath et al. reference. Hence, the teachings of Colbath et al. do not include or suggest generating a keyword index as in the present invention.

Also, as noted above, Colbath et al. does not teach a mechanism for associating a rule to a topic, as required by claim 1. The Examiner argues that col. 5, lines 1–33, of Colbath et al. teach this limitation. However, this is in error because col. 5, lines 1–33, of Colbath et al. teach (generally known) methods of speech recognition. The present invention, and in particular the third element of claim 1, is not concerned with speech recognition (although it may be compatible with speech recognition). The third element of claim 1 requires that each topic in the topic hierarchy is associated with a set of n-grams which are distinctive of that topic, so that searches can be performed.

Regarding claim 3, the Examiner argues that Colbath et al. teach a keyword index, and that reviewing the keyword index is also taught by Colbath et al. However, Colbath et al. do not teach a keyword index according to the present invention. Col. 2, lines 20–35, of Colbath et al., identified by the Examiner with

reference to claim 3, teaches that key words are searched for by providing them to a search engine. Col. 2, lines 20–35, does not teach a keyword index as in the present invention, wherein the keyword index is created from Web pages and is a field searched in. Hence, Colbath et al. do not meet the limitations of claim 3.

Claims 2, 4 and 5 were rejected under 35 U.S.C. §103(a) as being unpatentable over the patent to Colbath et al. in view of U.S. Patent No. 5,819,220 to Sarukkai et al. This rejection is respectfully traversed for the reason that the combination of Colbath et al. and Sarukkai et al. does not fairly teach or suggest the claimed invention.

Regarding claim 4, Colbath et al. do not teach “creating rules from the sparse n-grams, wherein each topic has associated rules that are used to decide if a new input document or query references the topic”, as amended. This is because Colbath et al. do not teach a natural language interface, and Colbath et al. do not teach that topics have associated rules. Colbath et al. teach only a voice activated search or web browser, as explained above. The above-quoted limitation from claim 4 requires that Web pages or documents be classified into a topic hierarchy so that they may be searched according to the present invention. Colbath et al. do not teach setting up topics or classifying data so that it can be searched, and hence do not meet this limitation of claim 4.

Sarukkai et al. do teach the use of n-gram language models. However, the teachings of Sarukkai et al. are not really applicable to the present invention because they are not directed toward the set-up of a natural language interface. Sarukkai et al. instead teach methods for dynamically altering language models according to word sets in the documents searched. In other words, the language model is adjusted in response to documents found in a search. The n-grams used by Sarukkai et al. are used for speech recognition, as known in the art. For example, Sarukkai et al. teach smoothing or re-estimating “n-gram *language model scores...*” (col. 9, lines 20–21, emphasis added), thereby implying that the n-grams are used for speech recognition. N-grams are extremely well known in the art. By comparison, the n-grams employed in the present invention are created

from documents to be searched, and the n-grams are stored as an index for searching. Hence, the n-grams in the present invention are used for very different purposes compared to the n-grams of Sarukkai et al. Consequently, the n-grams of Sarukkai et al. cannot reasonably be combined with Colbath et al. to meet the limitations of claims 2 or 4, as the Examiner argues.

The references cited by the Examiner and not relied upon or commented on have been reviewed; however, none of these references are believed pertinent to the claimed invention.

In view of the foregoing, it is respectfully requested that the application be reconsidered, that claims 1 to 6 be allowed, and that the application be passed to issue.

Should the Examiner find the application to be other than in condition for allowance, the Examiner is requested to contact the undersigned at the local telephone number listed below to discuss any other changes deemed necessary in a telephonic or personal interview.

A provisional petition is hereby made for any extension of time necessary for the continued pendency during the life of this application. Please charge any fees for such provisional petition and any deficiencies in fees and credit any overpayment of fees to Attorney's Deposit Account No. 50-2041.

Respectfully submitted,



C. Lamont Whitham  
Reg. No. 22,424

Whitham, Curtis & Christofferson, P.C.  
11491 Sunset Hills Road, Suite 340  
Reston, VA 20190  
Tel. (703) 787-9400 or (703) 391-2510  
Fax. (703) 787-7557 or (703) 391-9035



30743

PATENT TRADEMARK OFFICE